



## THOR: Metrics and Tools

### Document Information

**Date:** 01/03/2016

**Author(s):** Robin Dasler (CERN)

**Reviewer(s):** Amir Aryani (ANDS)  
Laura Rueda (DataCite)

**Abstract:** This report describes the work of the THOR Project to develop a dashboard to monitor interoperability of persistent identifiers. The dashboard is an essential step towards a suite of tools to measure the impact of the project.

**DOI** 10.5281/zenodo.46761

This work was supported by the THOR Project. The THOR project is funded by the European Union under H2020-EINFRA-2014-2 (Grant Agreement number 654039). The following report is based on a deliverable submitted to the European Union on 1 December 2015.

Visit <http://project-thor.eu> for more information.



## Contents

<b>Executive summary</b>	<b>1</b>
<b>1 Introduction</b>	<b>1</b>
1.1 <i>Dashboard Resources at a Glance</i>	2
1.2 <i>Dashboard at a Glance</i>	2
<b>2 Scope and Orientation within THOR</b>	<b>3</b>
2.1 <i>Sustainability and the Iterative Feedback Loop</i>	3
2.2 <i>Quantitative vs. Qualitative Metrics</i>	3
<b>3 Planning and Prioritisation</b>	<b>4</b>
3.1 <i>Requirement Gathering</i>	4
3.1.1 Internal	4
3.1.2 External Feedback (Focus Groups)	4
3.2 <i>Use Cases</i>	6
3.2.1 Internal Use by THOR Project Team and THOR Partners	6
3.2.2 Data Centre Metadata Enhancement	6
3.2.3 Assessing Wider Impacts of Funding	6
3.2.4 Demonstrating the Impacts of Open Data	7
3.3 <i>Priorities for Immediate Inclusion</i>	8
3.3.1 ORCID and DataCite as Baseline Metrics	8
3.3.2 Evaluation of Outreach Activities	9
3.3.3 Funding Information	9
3.3.4 Transparency of Underlying Data	9
3.4 <i>Challenges</i>	10
3.4.1 Provision of Data to Feed the Dashboard	10
3.4.2 Institutional Statistics	11
3.4.3 Data Citation Statistics	11
<b>4 Implementation</b>	<b>11</b>
4.1 <i>Development Philosophy</i>	11
4.2 <i>Metrics Included</i>	12
4.2.1 Researcher Identifier Metrics (ORCID)	12
4.2.2 Research Object Identifier Metrics (DataCite)	13
4.2.3 Outreach Metrics	15
4.3 <i>Interface Design</i>	15
<b>5 Future Work</b>	<b>19</b>
5.1 <i>Future Dashboard Versions</i>	19
5.2 <i>Dashboard Sustainability</i>	19
<b>6 Conclusion</b>	<b>19</b>
<b>Appendix A: Terminology</b>	<b>20</b>
<b>Appendix B: Project summary</b>	<b>21</b>



## Executive summary

As part of project sustainability efforts, THOR was charged with developing a suite of measurement tools with which to monitor the evolution of persistent identifier interoperability and to evaluate the impact of the project with influencing that evolution. In response, we have developed a metrics dashboard, building initially on the statistics available from our partners ORCID and DataCite as representatives of researcher and research object metrics, respectively. This dashboard is developed using common Open Source tools and is readily extendable to include other metrics in the future. In order to keep pace with the evolving nature of the services offered by our partners, the dashboard will see additional revisions over the life of the THOR project.

In the process of developing the dashboard, the sustainability team also fed the work of research and service development within THOR by identifying which information was readily available for harvest, which additional information was desired by stakeholders, and where more work was needed to make the inclusion of such information feasible.

## 1 Introduction

The THOR project seeks to improve the interoperability of persistent identifiers (PIDs) by fostering the infrastructure necessary to make the continued development of interoperable systems a reality. The growth of this infrastructure, both in terms of technical capabilities and awareness on the part of research and information stakeholders, will be the primary measure of THOR's success.

In light of this, the THOR project is developing a suite of measurement tools with which to publicly evaluate the success of both persistent identifier interoperability as well as the project itself. This has been realised by implementing a metrics dashboard, a real time web-based depiction of quantifiable metrics related to the project and the overall global development of persistent identifiers. This dashboard is publicly available (see Table 1), in keeping with THOR's overall theme of building community engagement around PIDs.

One of the most important contributions of this work lies in what is not currently visible in the dashboard. A major component of determining how to measure progress was first determining which measures were available, as it is impossible to aggregate information that does not exist. For the automation of the dashboard to be successful, data relevant to THOR's sustainability purposes needed both to exist and to be readily harvestable. Where these data were non-existent or not available, steps were taken to either make them available or to determine the feasibility of their future availability. This work has already led to improvements in the data available via our partners, described in more detail below (see Section 3.4.1).

This is the first report on the dashboard and the first iteration of the dashboard, giving the THOR project a foundation from which to build and evaluate over the life of the project. The dashboard will continue to evolve over time as additional interoperability capabilities are realised through the work of the THOR work packages. These later updates will be documented as part of regular project reviews.



## 1.1 Dashboard Resources at a Glance

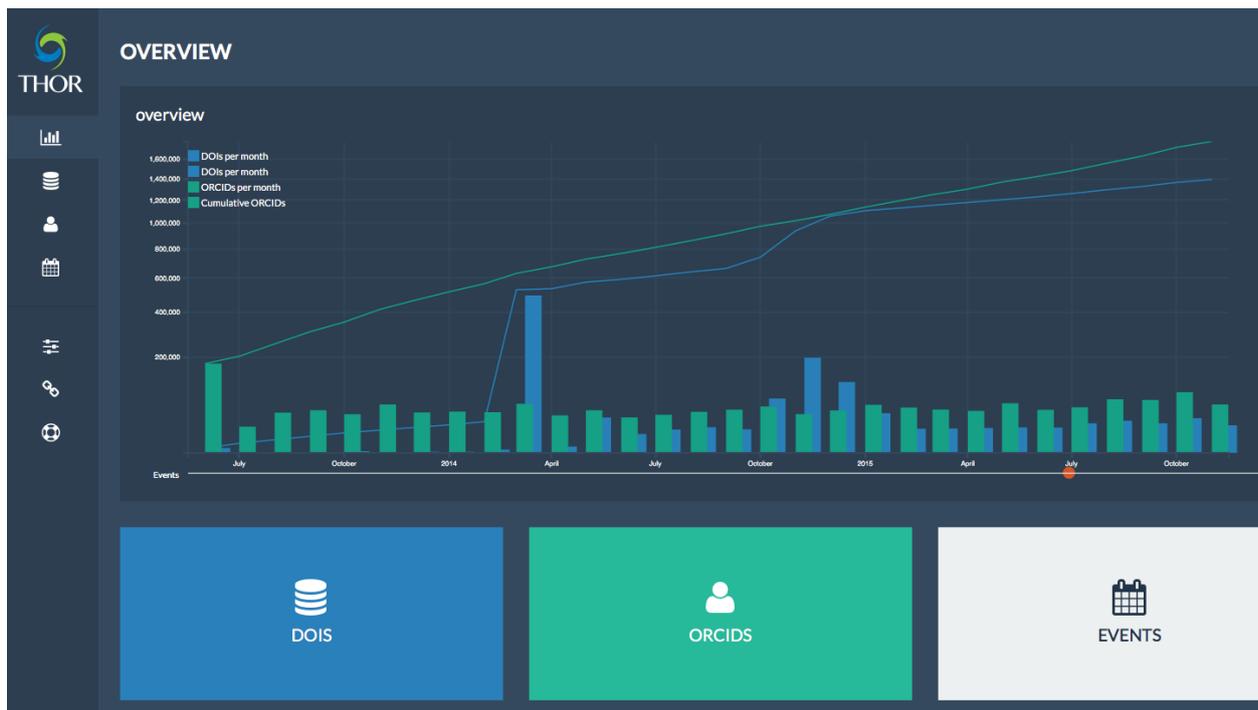
Table 1: Publicly Available Dashboard Resources

Resource	URL
Dashboard	<a href="http://dashboard.project-thor.eu">http://dashboard.project-thor.eu</a>
Dashboard code repositories	<a href="http://github.com/thor-project/dashboard">http://github.com/thor-project/dashboard</a> <a href="http://github.com/thor-project/data-harvester">http://github.com/thor-project/data-harvester</a>
API endpoint, ORCID Statistics	<a href="https://pub.orcid.org/v2.0_rc1/statistics">https://pub.orcid.org/v2.0_rc1/statistics</a>
API endpoint, DataCite Metadata Search	<a href="http://search.datacite.org/api">http://search.datacite.org/api</a>

## 1.2 Dashboard at a Glance

A screenshot of the dashboard homepage is below (see Figure 1). Additional screenshots and detailed descriptions of dashboard features can be found in Section 4.3.

Figure 1: Metrics Dashboard Overview Tab





## 2 Scope and Orientation within THOR

### 2.1 Sustainability and the Iterative Feedback Loop

The role of the THOR sustainability team is to investigate the long-term sustainability of THOR's efforts and provide an iterative feedback cycle for informing and validating the work of the project's other research and implementation outputs. A major component of initiating the work of the metrics dashboard was determining which information about the persistent identifier landscape was most pertinent to the THOR project and which information was currently available for harvest and aggregation. Already this first step of requirement engineering led to a number of improvements in the data available from partners (see Section 3.4.1) and identification of priorities for future work of THOR overall.

Furthermore, as the intention of the dashboard was to provide a framework for informing other work packages, the influence of the dashboard will be seen in project reporting, particularly for the THOR sustainability deliverables.

### 2.2 Quantitative vs. Qualitative Metrics

As the task at hand was to develop measurement tools, the resulting product consists of those metrics that could be quantified. This naturally leaves some gaps in describing the full picture of persistent identifiers and their use and importance in research workflows, as well as understanding the impact of THOR outputs. The quantitative metrics presented in the dashboard can be used as a starting point for conducting broader qualitative assessment of these topics. The dashboard may also assist in driving additional areas of qualitative research, so that THOR may get the complete portrait of researcher and institution motivations and cooperation in improving PID interoperability.

In particular, there are several guiding questions presented in the original description of work in regards to sustainability for that require more than a purely quantitative answer. These are:

- **Disciplinary data platforms:** Who has integrated ORCIDs? DOIs? Other PIDs? Which platforms provide users citation structure for submitted datasets? Which platforms support collection and/or connection between identifier types?
- **Scholarly communications:** Which stakeholders and organisations are using PIDs in research workflows? How? Are the PID linkages created in these workflows propagated or leveraged by other systems?
- **Researchers:** Do researchers know about PIDs? How many researchers sign up for ORCID following encouragement by their university or association? Do researchers have the opportunity to use PIDs during the research data cycle? Do researchers use PIDs to cite data/link data to publications?



### 3 Planning and Prioritisation

The dashboard will meet the goals outlined in the description of work for the sustainability arm of THOR (see Table 2). In order to determine the specifics of how best to achieve these goals, we went through a multistage process of gathering requirements, documenting use cases, and prioritizing development work.

Table 2: Dashboard Goals

Task	Goal
1	Monitor use of PIDs for data sharing, citation, and re-use
2	Assess PID integrations in research workflows across stakeholders and disciplines
3	Monitor active use and response to project tools, services, and outreach

#### 3.1 Requirement Gathering

The requirement gathering process (see Figure 2) for the metrics dashboard began internally, as the dashboard is primarily a tool to monitor and showcase progress toward project goals. A first prototype was developed from those requirements, and this prototype was presented in the focus groups four months into the project for vetting by experts (see Section 3.1.2).

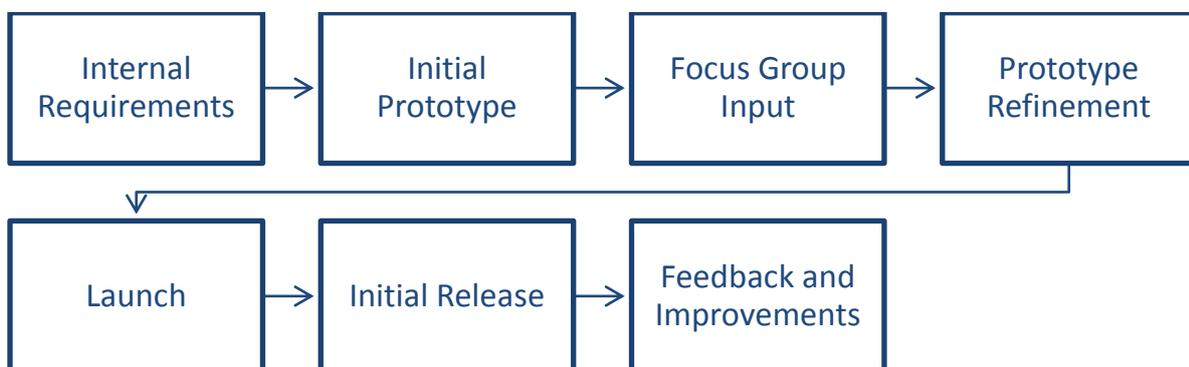


Figure 2: Requirement Gathering Process for THOR Metrics Dashboard

##### 3.1.1 Internal

Initial requirements gathering was conducted internally with the THOR partners. Priorities were first established at the project kick-off meeting, and the details of definitions and available data were determined through conversations with the relevant technical leads. Through this process, the first outline of proposed metrics was developed. The primary value in this process was in identifying the gaps in the data currently harvestable from our project partners and from the broader identifier landscape.

##### 3.1.2 External Feedback (Focus Groups)

It is important to gather expertise from international experts, offering insights into their needs and expectations for such a tool. Focus groups to gather input and feedback on the proposed THOR metrics dashboard were therefore held immediately before the Research Data Alliance Sixth Plenary in Paris at the end of September 2015. Twelve experts representing our identified community stakeholders took



part. These participants were hand selected based on their involvement in the research data community and their expertise in working with or developing identifier systems of their own (see Table 3). Having representation from academic institutions, service providers, and publishers, among others, meant that our focus groups benefited from a wealth of specific expertise from the very people who are potential users of our tool.

Table 3: Focus Group Participant Profiles

Stakeholder Category	Country
Academic Institution/Consortium	Netherlands
Academic Institution/Consortium	Switzerland
Academic Institution/Consortium	UK
Academic Institution/Consortium	USA
Academic Institution/Consortium	USA
Funder	Belgium
IT Infrastructure Provider	Germany
IT Infrastructure Provider	Greece
Research Producing Organisation	France
Research Producing Organisation	France
Scholarly Publisher	Netherlands
Service Provider	USA

During the focus groups, participants first engaged in a discussion of the kinds of statistics they would most like to see, and how they might use those statistics to make decisions in their own work with identifiers. Then participants were shown an early prototype of the dashboard, featuring a proposed layout and generated imaginary data, and asked to compare this prototype with the expectations they had generated. This agenda helped to elicit an independent, unbiased understanding of stakeholder expectations first, which could then be compared to the THOR developments and prototype to determine whether those expectations were met, resulting in more constructive feedback.

In summary, the results of the focus groups provided excellent feedback and insights for the dashboard and for the THOR project as a whole. The participants unanimously supported the current design, use cases, architecture, and content of the dashboard. The focus groups suggested feature requests and validated planned operations of the THOR dashboard, and this influence can be seen throughout the remainder of this paper as well as in the finished dashboard product. Additionally, focus group participants were eager to be involved in a second round of feedback for a later version of the dashboard. While not all feature requests are able to be implemented in the iteration of the dashboard that accompanies this paper, these requests have provided guidance for future work in research, services, and outreach for THOR to address the current gaps observed (thanks to THOR's iterative feedback model). In particular, the focus groups helped to improve and refine our set of use cases for the dashboard.



## 3.2 Use Cases

The dashboard is envisioned for use by a number of key stakeholder groups. It should be noted that the dashboard is primarily intended as an aggregator of high-level persistent identifier trends across the research landscape, and our key stakeholder groups are therefore primarily comprised of institutional decision makers. Individual researchers would be better suited to turn to specific identifier providers servicing their disciplines for metrics related to their specific outputs.

In all cases, the dashboard will provide an overall picture of project impact and progress. The range of disparate use cases will require the dashboard to be easy to understand and manipulate by people from different backgrounds and with different goals.

The focus groups were of great help in defining use cases for the dashboard. Three specific use cases, aside from the primary internal use case, are derived from the discussion in the focus groups directly. Specific measures relevant to the THOR project's strategic goals can be mapped to each of the use cases (see Table 4).

### 3.2.1 Internal Use by THOR Project Team and THOR Partners

In addition to the community stakeholders, the dashboard will serve as a primary component of the iterative feedback loop that will inform the work of the THOR project team. The dashboard will allow the project team to view an overall picture of project progress, allowing the team to make value judgments in regards to project effort and assisting with subsequent evaluation and reporting. The collection of PID interoperability metrics will be largely automated, benefiting from the information the THOR partners make publicly available. These metrics will inform our progress toward harmonisation of technical infrastructure. For the human infrastructure perspective, automating gathering of outreach and training statistics was currently impractical. Instead, the dashboard includes an outreach overlay fed by a manual entry form (described in Section 4.3) that enables the team to record outreach statistics.

### 3.2.2 Data Centre Metadata Enhancement

One use case arising from the focus groups was centred around the participants' desire to use the information from the dashboard to facilitate their role as providers of quality metadata. Seeing aggregate information about the metadata provided for deposited datasets could help data centres determine how well the metadata they deposit aligns with the aggregated norm. By giving the data centres a means to see how their metadata contributes to a greater whole, they could be encouraged to provide more consistent and higher quality metadata, thus promoting a circle of metadata enhancement between data centres and identifier organisations. Focus group participants were eager to close this circle and take an active role in improving metadata.

### 3.2.3 Assessing Wider Impacts of Funding

Greater interoperability between PIDs provides an opportunity to more thoroughly assess the impacts of research funding by allowing funders and others to trace the network of influence that connects researchers, data, and funding. These relationships currently are not easily surfaced. The addition in spring 2016 of metadata linking researchers and datasets to funding agencies following a planned extension of the DataCite Metadata Schema will allow the dashboard to include this vital piece of the



research cycle in a view of the whole. While this will be a welcome addition for many in our stakeholder community, it will be especially useful to funding agencies, allowing them a way to monitor the wider impacts of their funding efforts in the broader context of the research ecosystem. The ability to observe the spread of research funding and its connection to multiple institutions and works, as well as its influence over the entire research landscape, will be a valuable indicator of funders' return on investment.

### 3.2.4 Demonstrating the Impacts of Open Data

Focus group feedback indicated that many of our stakeholders have a desire to demonstrate the impacts of open data to others both inside and outside their institutions, both for policy compliance and for researcher support. Having the dashboard as a tool to monitor the connections between data, researchers, scholarly works, and funding will be a simple "one stop shop" option for these stakeholders to observe and disseminate the growing network of interoperable identifiers in a visually impactful way. To this end, focus group participants were particularly interested in the ability to view statistics on researchers and datasets that pertained to their individual institutions. Unfortunately, this is not feasible for either the dashboard or the THOR project partners at this time. Both institutional statistics and data citation statistics remain open challenges (see Sections 3.4.2 and 3.4.3 respectively), and have reinforced the priorities for exploring institutional identifiers and data event tracking that are emerging as part of THOR's research efforts.

Table 4: Alignment of Measures to Use Cases

Goals & Measures		Relevant to Use Cases			
		Internal Use by THOR and Partners	Data Centre Metadata Enhancement	Assessing Wider Impacts of Funding	Demonstrating Impacts of Open Data
1	Monitor use of PIDs for data sharing, citation, and re-use				
1.1	Baseline counts				
	ORCID presence	X			
	DataCite presence	X			
1.2	Interconnections				
	From authors to works	X	X		X
	From authors to institutions	X	X		X
	From authors to funders	X	X	X	X
	From data to authors	X	X		X
	From data to institutions	X	X		X
	From data to funders	X	X	X	X
1.3	Citation counts				



Goals & Measures		Relevant to Use Cases			
	For data sets	X	X	X	X
	For authors	X	X	X	X
2	Assess PID integrations in research workflows across stakeholders and disciplines				
2.1	Stakeholder reach				
	Object types represented	X	X		X
	Institutions represented	X	X	X	X
	Funders represented	X	X	X	X
2.2	Disciplinary reach				
	Disciplines represented	X	X	X	X
3	Monitor active use and response to project tools, services, and outreach				
3.1	Impact of project tools and services				
	Website traffic	X			
	Social media engagement	X			
	Documentation engagement	X			
3.2	Impact of project outreach events				
	Event attendance	X			
	Event reach	X			
	Event influence on project tools and services	X			

### 3.3 Priorities for Immediate Inclusion

Through the requirements, focus group, and feedback gathering processes, a number of features and components were identified and subsequently prioritised according to feasibility of implementation. The following are the priorities that informed development of this iteration of the dashboard.

#### 3.3.1 ORCID and DataCite as Baseline Metrics

The THOR sustainability team are responsible for developing metrics and key performance indicators by building on baseline statistics of the existing persistent identifier infrastructure as exemplified by the ORCID and DataCite services. The dashboard being delivered therefore relies on ORCID as the exemplar of researcher identifier metrics and on DataCite as the data identifier exemplar, laying the foundation for a metrics framework that can later be expanded to include other identifiers. The dashboard provides basic statistics on both of these services over time, including counts of registered identifiers and links between the identifier types. The dashboard data harvester module (see Section 4.1) can be easily



updated to add additional statistics from other providers, assuming there is an available API endpoint that provides flat statistics.

### 3.3.2 Evaluation of Outreach Activities

THOR is required to evaluate the use of and response to project services, tools, training materials, and documentation. The metrics dashboard will therefore include a component to track engagement with project deliverables, events, programmes, and communication streams. For this iteration of the dashboard, an outreach events overlay (described in more detail in Section 0) is fed from a manual entry form, collecting basic attendance and demographic information from events. Further standardised metrics related to engagement will be finalised and implemented at the beginning of 2016. In keeping with the automated harvesting in the rest of the dashboard, the goal is to gather this information from the external tools we are using to manage our outreach materials, including:

- Twitter (followers/mentions);
- readme.io (project documentation interaction);
- Google Forms (event registration/demographics).

### 3.3.3 Funding Information

Focus group participants requested the addition of funding information as a critical component in monitoring the persistent identifier ecosystem. Specifically, they were interested to know which data sets were funded by which funders, and were hopeful that a tool like the THOR dashboard could possibly show this information in the future. While determining funding information down to the individual grant is essentially impossible at this time, this iteration of the dashboard does include some high level general statistics about funding information included in ORCID records. Further funding statistics for data sets will be incorporated following the release of the DataCite Metadata Schema 4.0. Linking authors and data to funding information is also an excellent example and use case for the project-level metrics proposed for investigation by THOR in the 'Artefact, Contributor, and Organisation Relationship Data Schema' report.<sup>1</sup>

### 3.3.4 Transparency of Underlying Data

Focus group participants saw a need for transparency in the data underlying the dashboard, both for their own purposes as decision makers and as a means of informing their clients, such as research managers, librarians, funders, and researchers, who are the basis for the data. There was a concern that metrics of this sort could be taken as a de facto ranking system or imply comparisons between disciplines that are not practical, and better transparency of data could help with these concerns.

This transparency is achieved through documentation available on the dashboard site, which was already planned, and through a number of minor improvements in the dashboard presentation that were requested during the focus groups. Furthermore, participants were interested in an API for

---

<sup>1</sup> Fenner, Martin, Demeranville, Tom, Kotarski, Rachael, Vision, Todd, Rueda, Laura, Dasler, Robin, ... THOR Consortium. (2015). *THOR: Artefact, Contributor, and Organisation Relationship Data Schema*. Zenodo. <http://doi.org/10.5281/zenodo.30799>.



accessing the statistics for their own projects and purposes. This is in keeping with the THOR project's emphasis on open scholarship and has already been planned for inclusion in early 2016.

### 3.4 Challenges

Through the process of gathering requirements, both internally and in the focus groups, a number of issues were identified that prompted additional work. This included the data that was available from partners (and its access), along with feature requests from stakeholders. A few of these issues remain as open challenges to both the development of the dashboard and the work of THOR overall.

#### 3.4.1 Provision of Data to Feed the Dashboard

At the outset of requirements gathering, not all of the data deemed necessary to monitoring the persistent identifier landscape was readily harvestable from the project partners. The various reasons for this included issues of metadata harmonisation (further explored in the *Artefact, Contributor, and Organisation Relationship Data Schema* report),<sup>2</sup> as well as impractical or inefficient data harvesting due to lack of an entry point for outsiders.

This resulted in metadata harmonisation work that is ongoing for THOR and is being informed by the focus group discussions led in support of dashboard development. The focus group participants consistently raised metadata quality as an area of concern, particularly in regards to completeness and harmonisation between schemes. Participants indicated a desire for guidance in improving the metadata their institutions provide (as indicated in Section 3.2.2). In particular, they indicated desires for (1) automated correction of submitted metadata; and (2) an indicator of metadata quality on a per record basis. Each of these is outside the scope of the dashboard task but could be addressed by THOR partners. In the first case, this is because authority for correction belongs to the individual identifier organisations. In the second, this is because a quantified indicator of metadata quality is beyond what is practical to expect from the dashboard. However, the expression of these interests reaffirms that consistent, reliable, and translatable metadata is at the heart of the identifier interoperability problem. Furthermore, it confirms that THOR's identified stakeholder community is conscious of and concerned about these problems of metadata and their scale.

In regards to data harvesting, it is necessary for the dashboard to have ready access to harvestable data related to counts and statistics at the identifier scheme level, without the need for frequently querying and parsing individual identifier records. This type of ready access was not initially available for ORCID metrics, but has subsequently been developed as part of a statistics API.<sup>3</sup> Similarly, DataCite is working on additional products and services, such as the DataCite Claim Store, that will allow for the harvesting of additional metrics for the dashboard to aggregate.

---

<sup>2</sup> Fenner, Martin, Demeranville, Tom, Kotarski, Rachael, Vision, Todd, Rueda, Laura, Dasler, Robin, ... THOR Consortium. (2015). *THOR: Artefact, Contributor, and Organisation Relationship Data Schema.* Zenodo. <http://doi.org/10.5281/zenodo.30799>.

<sup>3</sup> ORCID Statistics API v2.0\_rc1. [https://pub.orcid.org/v2.0\\_rc1#!/Statistics\\_API\\_v2.0\\_rc1/viewStatsTimeline](https://pub.orcid.org/v2.0_rc1#!/Statistics_API_v2.0_rc1/viewStatsTimeline).



### 3.4.2 Institutional Statistics

The dashboard was intended to monitor the evolving interoperability of persistent identifiers at a high level. In the focus groups, participants indicated that they would value the ability to easily view statistics for authorship and data for their specific institution. This level of granularity is not currently possible given the available tools and available information sources, but it could be feasible given some additional work on the part of our project partners and others in the identifier ecosystem. Additional work on solidifying institutional identifier use and practices is necessary, which is already in discussion as part of THOR's research efforts.

### 3.4.3 Data Citation Statistics

Focus group participants were particularly interested in a tool that would provide statistics around data citation, in order to demonstrate and/or make decisions around the impact of open data. This is another highly valued request that is currently impractical given the tools and information available for us to harvest, but it is also feasible given work by our partners. In particular, the DataCite Data Level Metrics<sup>4</sup> service is already a work in progress. The THOR metrics dashboard will be able to harvest data citation event information via this tool beginning in 2016.

## 4 Implementation

The THOR metrics dashboard is being implemented as a lightweight harvester and aggregator of data existing from other sources. The goal is to provide a framework for monitoring and developing an interoperable persistent identifier ecosystem, rather than inventing specific add-on identifier services.

The dashboard development pursues quality by adhering to principles of openness, flexibility, portability, and simplicity. This will ensure the product's sustainability and longevity by making it self-explanatory in design and easily mobile.

### 4.1 Development Philosophy

In keeping with THOR's commitment to Open Science, the metrics dashboard is being developed with an open philosophy. Code is publicly available on GitHub<sup>5</sup> under a BSD license.

The dashboard is being developed with common non-proprietary tools, primarily Python<sup>6</sup> and D3.js<sup>7</sup>. This will aid in easy portability to other systems while furthering our goals of openness.

The dashboard is developed in two modules: "data-harvester"<sup>8</sup> to ingest data from defined sources and flatten it into a form that can be used; and "dashboard",<sup>9</sup> which contains the tools for visualising the

---

<sup>4</sup> <http://dlm.labs.datacite.org/>

<sup>5</sup> <http://github.com/thor-project>

<sup>6</sup> <https://www.python.org/>

<sup>7</sup> <http://d3js.org/>

<sup>8</sup> <https://github.com/thor-project/data-harvester>

<sup>9</sup> <https://github.com/thor-project/dashboard>



data. Additional sources can be added to the data-harvester module easily, assuming they provide an API endpoint for harvesting statistics, allowing for future expansion to include other identifier systems.

## 4.2 Metrics Included

### 4.2.1 Researcher Identifier Metrics (ORCID)

For this version, the metrics harvested from ORCID will largely be limited to the metrics available via the recently launched ORCID statistics API<sup>10</sup> (see Table 5), which also powers ORCID’s own statistics page.<sup>11</sup>

Metrics available via ORCID’s statistics API cover the entire body of ORCID’s registered, while the public API handles calls for information about particular ORCID’s. Since THOR is primarily concerned with metrics around the landscape of persistent identifiers at a high level, the dashboard will harvest information from the statistics API, which is also a more efficient practice than collecting information on many individual ORCID’s. This means the dashboard will be able to show statistics on the number of works with DOIs claimed across the entire ORCID system or the number of claimed employment affiliations across the system, but breakdowns of individual types of works or individual researcher associations are currently not available to harvest efficiently.

Table 5: Statistics Gathered from ORCID Statistics API

Goal	Property Definition	Measure	Source (Query Field)
1	Monitor use of PIDs for data sharing, citation, and re-use		
1.1	Baseline counts		
	ORCID presence	Total number of ORCID identifiers registered monthly	Query <code>liveIds</code> for running total. Take difference between monthly values to plot change.
1.2	Interconnections		
	From authors to works	Total number of ORCID records that include works	Query <code>idsWithWorks</code> for running total. Take difference between monthly values to plot change.
		Total number of works contained in all ORCID records	Query <code>works</code> for running total. Take difference between monthly values to plot change.

<sup>10</sup> API endpoint: [https://pub.orcid.org/v2.0\\_rc1/statistics](https://pub.orcid.org/v2.0_rc1/statistics)

<sup>11</sup> <http://orcid.org/statistics>



Goal	Property Definition	Measure	Source (Query Field)
		Total number of unique DOIs that exist across all ORCID records	Query uniqueDOIs for running total. Take difference between monthly values to plot change.
	From authors to institutions	Total number of ORCID records that have claimed institutional affiliation <ul style="list-style-type: none"> <li>• as employment</li> <li>• as education</li> </ul>	Query field employment for running total. Query education for running total. Take differences between monthly values to plot change.
	From authors to funders	Total number of ORCID records that include funding information	Query funding for running total. Take difference between monthly values to plot change.
2	Assess PID integrations in research workflows across stakeholders and disciplines		
2.1	Stakeholder reach		
	Institutions represented	Total number of institutions represented in ORCID (that can be uniquely identified) <ul style="list-style-type: none"> <li>• as employment</li> <li>• as education</li> </ul>	Query employmentUniqueOrg for running total. Query educationUniqueOrg for running total. Take differences between monthly values to plot change.
	Funding agencies represented	Funding agencies represented in ORCID (that can be uniquely identified)	Query fundingUniqueOrg for running total. Take differences between monthly values to plot change.

#### 4.2.2 Research Object Identifier Metrics (DataCite)

The majority of the metrics required can be accessed through the Solr-based DataCite Metadata Search API<sup>12</sup> (see Table 6), which also powers DataCite's Metadata Search<sup>13</sup>.

<sup>12</sup> API endpoint: <http://search.datacite.org/api>.

<sup>13</sup> <http://search.datacite.org>.



Table 6: Statistics Gathered from DataCite Metadata Search API

Goal	Property Definition	Measure	Source (Query Field)
1	Monitor use of PIDs for data sharing, citation, and re-use		
1.1	Baseline counts		
	DataCite presence	Number of DataCite identifiers minted	Query <code>minted</code> with current month date range
1.2	Interconnections		
	From data to authors	Total number of DataCite records that include an ORCID	Query <code>nameIdentifier for ORCID: *</code> and display total count
1.3	Citation counts [not currently feasible]		
2	Assess PID integrations in research workflows across stakeholders and disciplines		
2.1	Stakeholder reach		
	Object types represented	Total DataCite DOI counts by resource type	Query <code>resourceTypeGeneral</code> for each controlled term <sup>14</sup>
	Institutions represented	Total DataCite DOI counts by data centre (uploader)	Query <code>datacentre</code> for each THOR data centre <sup>15</sup> <ul style="list-style-type: none"> <li>• ANDS [allocator]</li> <li>• BL [allocator]</li> <li>• CERN [allocator]</li> <li>• DATACITE [allocator]</li> <li>• CDL.DRYAD [data centre]</li> <li>• TIB.PANGAEA [data centre]</li> </ul>

<sup>14</sup> DataCite Metadata Schema v. 3.1, Appendix 1: Controlled List Definitions, Table 7: Description of `resourceTypeGeneral`, p. 26. <http://dx.doi.org/10.5438/0010>.

<sup>15</sup> Elsevier, EMBL-EBI, ORCID, and PLOS are not listed data centres.



### 4.2.3 Outreach Metrics

Statistics related to outreach will be gathered manually through a form on the Events tab of the dashboard. For this release, the statistics gathered are as below (see Table 7).

Table 7: Statistics Gathered from Manual Entry Form

Goal	Property Definition	Measure	Source (Query Field)
3	Monitor active use and response to project tools, services, and outreach		
3.2	Impact of project outreach events		
	Event attendance	Total number of participants	Form entry under Participant Count
	Event reach	Total number of participants from each stakeholder group	Form entry under Audience Type
	Event influence on project tools and services	Plot event instances by date against Overview timeline to observe correlation	Form entry under Event Date

## 4.3 Interface Design

The total quantity of each identifier unit will be shown over time as a column chart, with a resolution of one month. Other associated statistics, such as counts per institution, will be shown as smaller secondary charts in blocks below the corresponding column chart (see Figure 3). Selection of an area of any chart will restrict the data shown based on the selection (see Figure 4). For example, selecting a time range in the primary chart will restrict the data in the secondary charts to only those data that correspond with that time range. Subsequent selection of a particular resource type in a secondary chart will further restrict the visible data to only those data that match both the selected resource type and the previously selected time range (see Figure 5).

Tabs exist for researcher metrics (see Figure 6) and for data set metrics (see Figure 3). These are currently populated by data from ORCID and DataCite, respectively. As mentioned previously, the kinds of statistics gathered, as well as the ease of updating the data-harvester module, will allow for future inclusion of other identifier metrics, assuming those metrics provide a means of harvesting their data.

The overview tab (see Figure 7) displays combined metrics for both researchers and data sets, as well as information relevant to the specific outreach efforts of the THOR project, which is plotted alongside the researcher and data set information according to date (note the orange dot in Figure 7). Selecting an outreach event will display more information about that event. For this version, the additional outreach information will come from manual form entry by the THOR project partners on the events tab (see Figure 8).

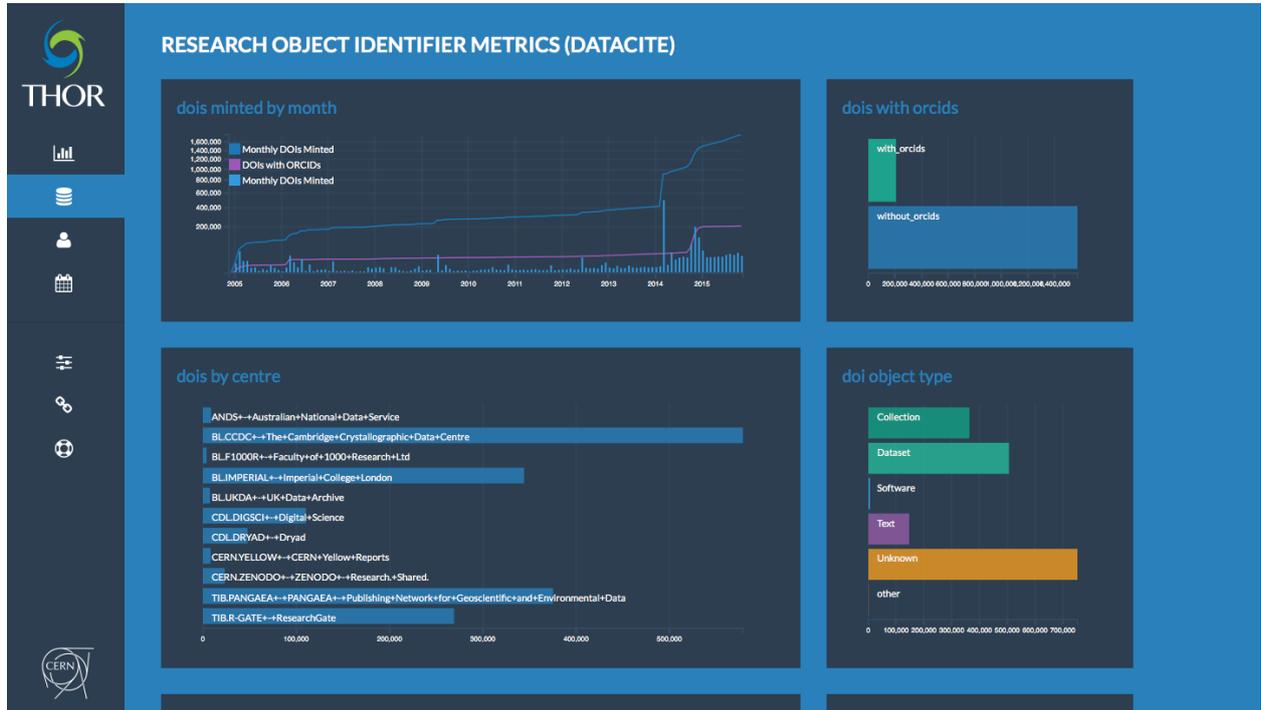


Figure 3: Metrics Dashboard Research Object Identifier Tab – Default View

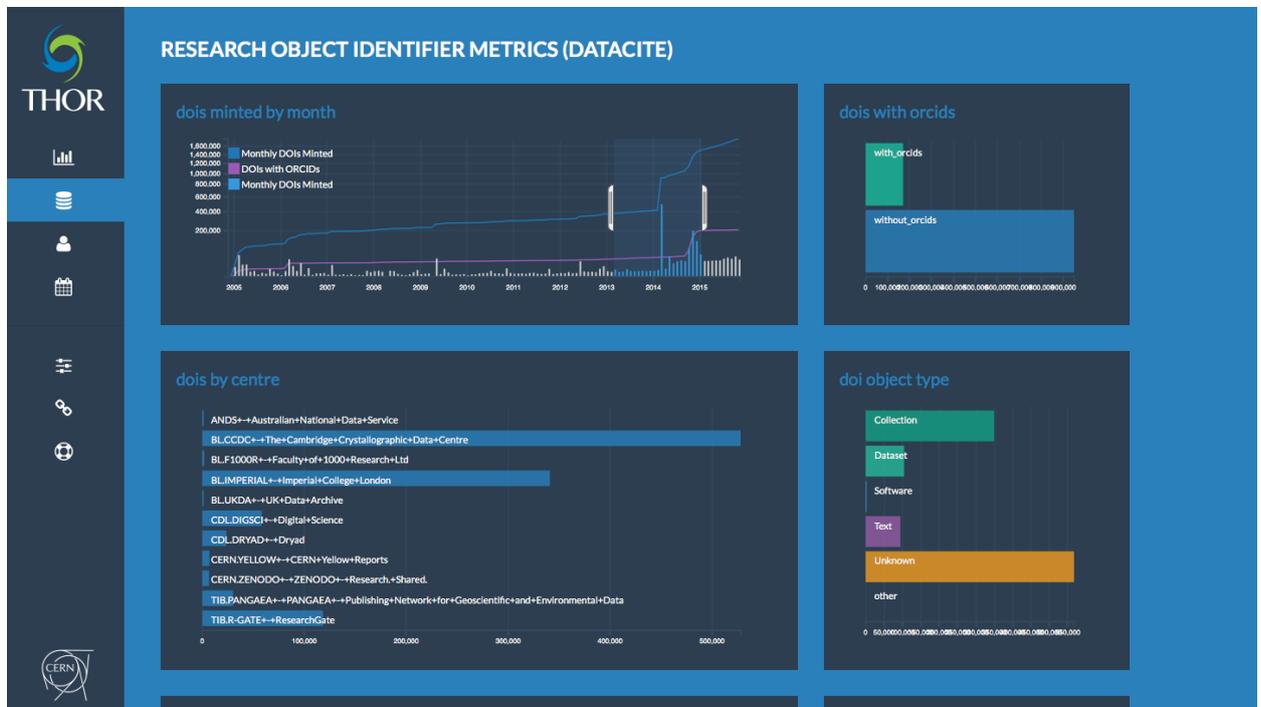


Figure 4: Metrics Dashboard Research Object Identifier Tab – Date Range Selection

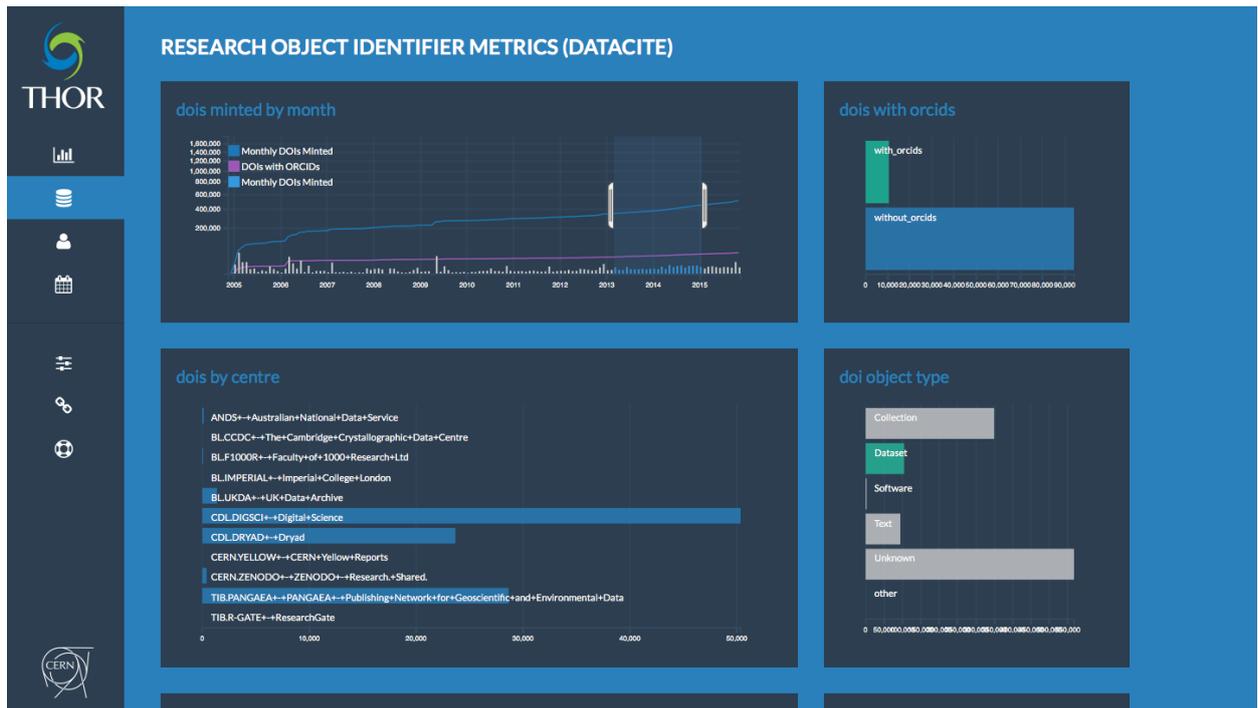


Figure 5: Metrics Dashboard Research Object Identifier Tab – Date Range and Object Type Selection

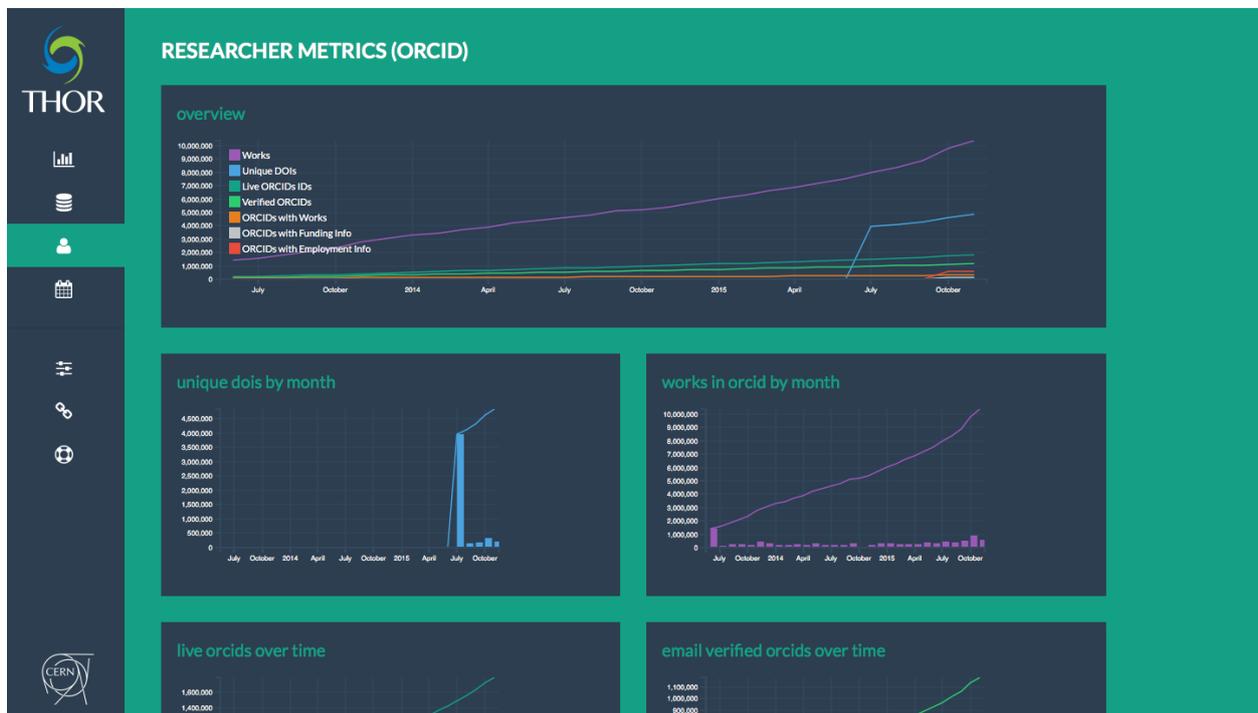


Figure 6: Metrics Dashboard Researcher Metrics Tab

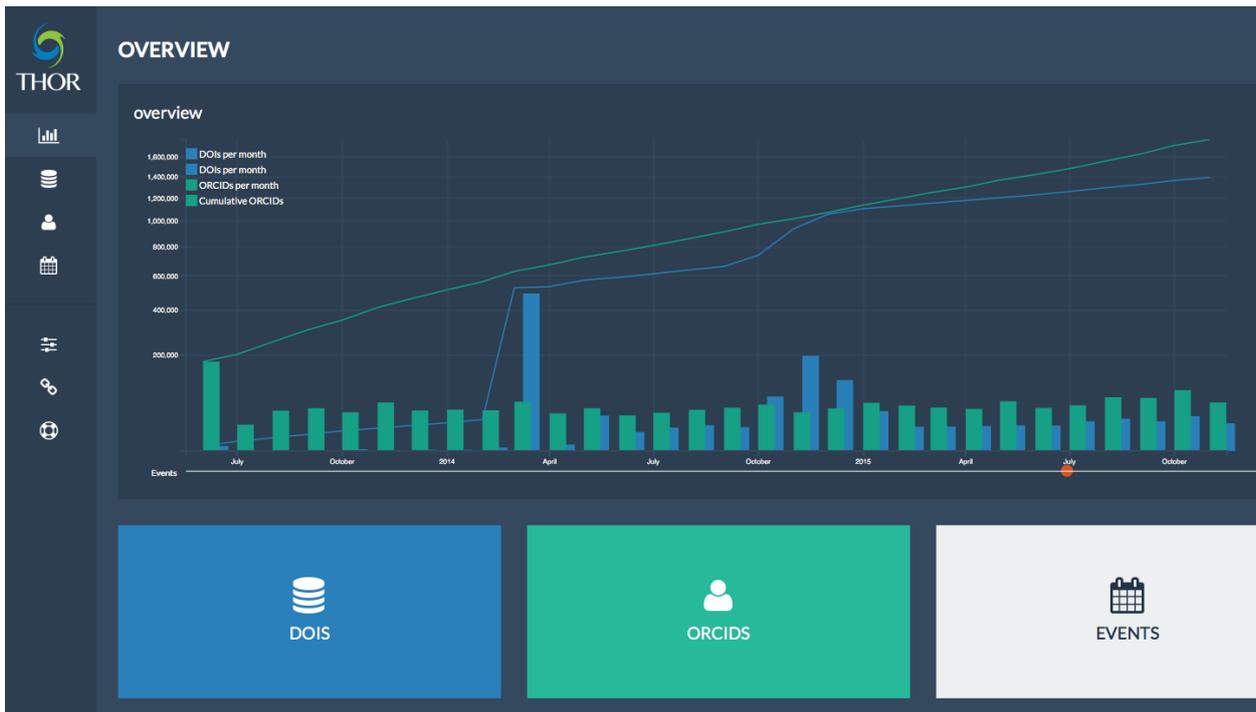


Figure 7: Metrics Dashboard Overview Tab

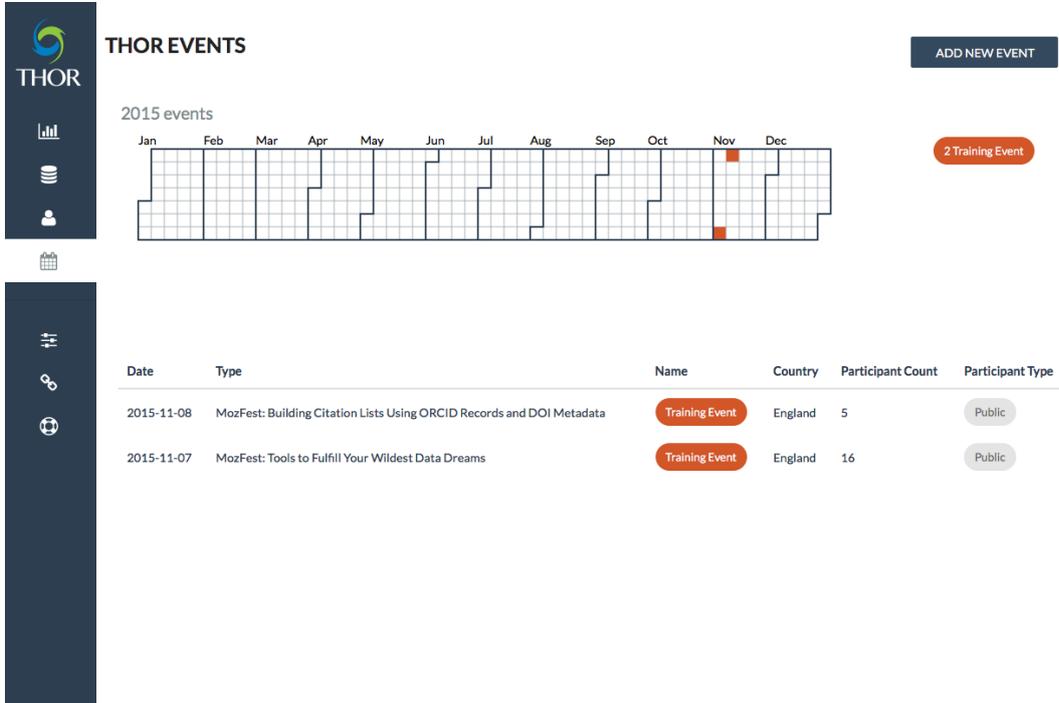


Figure 8: Metrics Dashboard Events Tab



## 5 Future Work

### 5.1 Future Dashboard Versions

This document describes an initial release of the dashboard, which will be updated and adapted over the course of the THOR project. Future updates will benefit from THOR's iterative feedback loop and will rely on further feedback and input from relevant sources. Future updates may vary based on other THOR outputs, but a general proposed roadmap follows:

- Initial public dashboard release (late 2015)
- Second round feedback from focus group participants (early 2016)
- Develop plan for incorporating metrics from social media, website analytics, and documentation hosting service (early 2016)
- Update data harvester to pull funding information from DataCite after release of DataCite Metadata Schema 4.0 (spring 2016)
- Investigate utility and feasibility of harvesting and including metrics from PIDs other than DOIs and ORCIDs (spring 2016)
- Develop API for accessing dashboard data (spring 2016)
- Update data harvester and dashboard interface to accommodate data citation information based on DataCite Claim Store release (spring 2016)
- Investigate feasibility of determining, providing, and harvesting information about disciplinary affiliation of data, works, and researchers (spring 2016)

### 5.2 Dashboard Sustainability

The dashboard is intended to last beyond the life of the THOR project. It has consequently been implemented without the use of uncommon proprietary frameworks, and it is designed to be readily and quickly portable. Its final home will be determined as part of the project's progress, but it is anticipated to be easily consumed as part of normal operations of one of THOR's project partners without much difficulty or overhead.

## 6 Conclusion

The metrics dashboard lays the foundation for the further work of the THOR project, particularly in regards to sustainability. The process of developing the tool has already helped identify gaps in existing service integration and provided further directions of work for other project teams within THOR. The expert focus groups provided valuable feedback and validation for our chosen directions and allowed us to generate interest and enthusiasm among our community stakeholders. The dashboard tool that has been developed based on these processes and community involvement is a robust and flexible mechanism for assessing the work and impacts of THOR.



## Appendix A: Terminology

Additional terms are defined below:

Term	Definition
DataCite	An organisation that develops and supports methods to locate, identify and cite data and other research objects. Specifically, DataCite develops and supports the standards behind persistent identifiers for data, and the members assign them. See <a href="https://www.datacite.org">https://www.datacite.org</a>
DOI	Digital Object Identifier
EC	European Commission
EU	European Union
ID	Identifier
ORCID	An organisation that creates and maintains a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers. See <a href="http://orcid.org">http://orcid.org</a>
PID	Persistent Identifier



## Appendix B: Project summary

The **THOR** project establishes a sustainable international e-infrastructure for persistent identifiers that enables long-term access to critical information about the life cycle of research projects. It enables seamless integration between articles, data, and researcher information creating a wealth of open resources. This will result in reduced duplication, economies of scale, richer research services, and opportunities for innovation.

The project has four concrete aims:

1. Establishing interoperability
2. Integrating services
3. Building capacity
4. Achieving sustainability

The project will meet these aims by defining relations between contributors, research artefacts (including data), and organisations. We will incorporate these relationships into the ORCID and DataCite systems. We will also expand existing linkages between different types of identifiers and versions of artefacts to improve interoperability across platforms and integrate ORCID iDs into production systems for article and data submission services in pilot communities and beyond.

The consortium will develop systems to embed new PID resolution techniques into existing services to support seamless direct access to artefacts, and in particular data. We will create services to allow associations between datasets, articles, contributors and organisations at the time of submission. Building on these, we will deliver the means to integrate trans-disciplinary PID services in community-specific platforms, focussing on cross-linking, claiming mechanisms and data citation (guided by the FORCE 11 data citation principles<sup>16</sup>).

For more information, visit <http://project-thor.eu> or contact [info@project-thor.eu](mailto:info@project-thor.eu)

---

<sup>16</sup> <https://www.force11.org/group/joint-declaration-data-citation-principles-final>